Inferência Estatística I

INTRODUÇÃO

TUANY CASTRO

Introdução

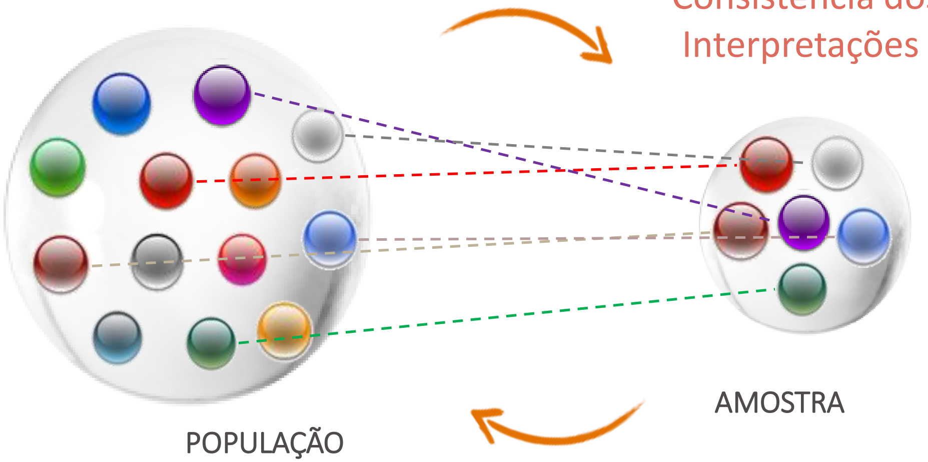
O uso de informações de uma amostra para concluir sobre o todo faz parte da atividade diária da maioria das pessoas. Basta observar como uma cozinheira verifica se o prato que ela está preparando tem ou não a quantidade adequada de sal. Ou, ainda, quando um comprador, após experimentar um pedaço de laranja numa banca de feira, decide se vai comprar ou não as laranjas. Essas são decisões baseadas em procedimentos amostrais.

A inferência estatística é um conjunto de técnicas que objetiva estudar a população através de evidências fornecidas por uma amostra.

Introdução

- **População**: conjunto de todos os elementos ou resultados sob investigação.
- **Amostra**: qualquer subconjunto da população.

Estatística Descritiva:
Consistência dos dados
Interpretações iniciais



Inferência Estatística:
Estimação de quantidades desconhecidas
Extrapolação dos resultados
Teste de hipóteses

Introdução

Exemplo 1: Suponha que estejamos interessados em estudar a proporção dos 1000 alunos, de uma escola do ensino médio, que pretendem fazer vestibular. Para tanto, selecionamos uma amostra de 20 alunos e perguntamos a eles sobre suas intenções futuras de estudo. Essas escolha pode ser feita de diversas maneiras:

- 20 alunos de uma mesma classe;
- 20 alunos espalhados entre as três séries;
- 10 meninas e 10 meninos;
- *Mais comum: associar um número a cada aluno e sortear 20 número aleatoriamente.*

Se você realize um sorteio aleatório dos alunos para a amostra e um amigo seu faça o mesmo procedimento. Você acha que as amostras sorteadas por você e seu amigo serão as mesmas? Podemos ter respostas próximas ou iguais nas diversas amostras?

Introdução

Conclusão:

- Não podemos garantir que repetições de amostras produzam sempre resultados idênticos;
- Todas as quantidades associadas à amostra têm caráter aleatório e, portanto, devem receber tratamento probabilístico;
- Se toda uma população entrar na amostra, temos toda a informação possível e portanto não há aleatoriedade.

Introdução

Exemplo 2: Seja X o número de caras obtidas depois de lançar uma moeda 50 vezes. Sabemos que X segue uma distribuição binomial com $n = 50$ e p a probabilidade desconhecida de se obter cara num lançamento (a moeda pode ser honesta ou não). Suponha que ocorrem 36 caras no lançamento da moeda. Esse resultado traz evidência de que a moeda seja honesta?

TESTE DE HIPÓTESES

1. P[illegible]rmos essa decisão, podemos partir do princípio de que a moeda não favorece nem cara nem coroa, isto é, $p = 1/2$. Com essa informação e com o modelo binomial, podemos encontrar a probabilidade de se obterem 36 caras ou mais e esse resultado nos ajudaria a tomar uma decisão.

ESTIMAÇÃO

2. Se [illegible]mos que a moeda não é honesta, qual é a melhor estimativa para p, baseando-se [illegible] amostra observada?

Introdução

Exemplo 3: Considere uma máquina de encher pacotes de café automaticamente. Digamos que ela esteja regulada para enchê-los segundo uma distribuição normal com média 500g e desvio-padrão 100g. Sabemos que, às vezes, a máquina desregula-se e, quando isso acontece, o único parâmetro que se altera é a média, permanecendo a mesma variância. Para manter a produção sob controle, vamos colher uma amostra de 100 pacotes e pesá-los. Como essa amostra nos ajudará a tomar uma decisão sobre a máquina?

- Parece razoável usarmos a média $\bar{x}$ da amostra como informação pertinente para uma decisão;
- Mesmo que a máquina esteja regulada, dificilmente $\bar{x}$ será igual a 500g, dado que os pacotes apresentam certa variabilidade no peso;
- Mas se $\bar{x}$ não se afastar muito de 500g, não existirão razões para suspeitarmos da qualidade do procedimento de produção. Só iremos pedir uma revisão se $\bar{x} - 500$, em valor absoluto, for muito grande.

Introdução

- O problema que se apresenta agora é o de decidir o que é próximo ou distante de 500g.
- Se pudéssemos colher várias amostras de 100 pacotes e observar $\bar{x}$ em cada uma delas, teríamos ideia do comportamento da média amostral e poderíamos dizer se o valor observado é ou não um evento raro.
- Portanto, é importante conhecer a distribuição da variável $\bar{x}$ para avaliar se está perto ou longe de 500g.

Introdução

Para formalizar as ideias que serão trabalhadas:

- **Parâmetro**: as quantidades da população, em geral desconhecidas, sobre as quais temos interesse. Usualmente representados por letras gregas tais como μ, σ, θ.
- **Estimador**: combinação de elementos da amostra, construída com a finalidade de representar ou estimar um parâmetro de interesse da população. Em geral, denotados por símbolos com acentos circunflexo: $\hat{\mu}, \hat{\sigma}, \hat{\theta}$.
- **Estimativa**: valores numéricos assumidos pelos estimadores.

Introdução

Exemplo 4: Estamos interessados na média das alturas de jovens com idade entre 15 e 18 anos, nascidos na região sudeste do país. Vamos coletar uma amostra e usá-la para tirar conclusões. Para facilitar a discussão, suponha que a amostra tenha somente 10 jovens, escolhidos ao acaso dentre a população mencionada.

- **População**: todos os jovens com idade entre 15 e 18 anos anos, nascidos na região sudeste.
- **Parâmetro de interesse**: altura média desses jovens, representada por μ.
- **Amostra** de tamanho 10: $(X_1, X_2, \dots, X_{10})$, observações das alturas dos jovens da amostra;
- **Estimador**: qual conta fazemos com as observações da amostra para estimar o parâmetro de interesse?

Introdução

- Possíveis estimadores:
  - $\hat{\mu}_1 = \frac{mínimo+máximo}{2}$;
  - $\hat{\mu}_2 = X_1$;
  - $\hat{\mu}_3 = \bar{X} = \frac{X_1+X_2+\cdots+X_{10}}{10}$

Suponha que a amostra coletada foi: 1,65; 1,57; 1,72; 1,66; 1,71; 1,74; 1,81; 1,68; 1,60 e 1,77. Então:

- Estimativas:
  - $\hat{\mu}_1 = \frac{mínimo+máximo}{2} = \frac{1,57+1,81}{2} = 1,69$;
  - $\hat{\mu}_2 = X_1 = 1,65$;
  - $\hat{\mu}_3 = \bar{X} = \frac{X_1+X_2+\cdots+X_{10}}{10} = \frac{1,65+1,57+\cdots+1,77}{10} = 1,69$

Qual estimador usar?

Introdução

Existem vários estimadores que podem ser calculados para estimar parâmetros de interesse. Porém, é interessante escolhermos aqueles que sejam mais assertivos. Para cada parâmetro, há um estimador mais assertivo:

Parâmetro	Estimador
μ: média populacional	$\bar{X} = \frac{X_1 + X_2 + \cdots X_n}{n}$
σ^2: variância populacional	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
p: proporção populacional	$\hat{p} = \frac{frequência\ amostral\ com\ a\ característica}{n}$

Exercícios

Exercício 1: Um fabricante deseja estudar a duração de baterias que são utilizadas em relógios de pulso. Uma amostra de vários lotes fabricados por uma mesma companhia foi submetida a testes acelerados e produziram os seguintes tempos de duração (em anos): 1.2,1.4, 1.7, 1.3, 1.2, 2.3, 2.0, 1.5, 1.8, 1.4, 1.6, 1.5, 1.7, 1.5 e 1.3. Determine estimativas para a média e a variância do tempo de duração dessas pilhas.

Exercício 2: O número de reclamações que chegam por hora a uma central de atendimento do consumidor foi anotado para uma amostra de algumas horas escolhidas ao acaso. Determine uma estimativa do número médio de reclamações e do desvio-padrão, se a amostra observada foi: 2, 2, 3, 1, 4, 5, 3, 4, 5, 6, 5, 3, 4, 3, 4, 2, 4, 3, 5, 2, 1, 6, 2, 3 e 4.

Exercícios

Exercício 3: Pretende-se estimar a proporção p de cura, através do uso de um certo medicamento em doentes contaminados com cercaria, que é uma das formas do verme da esquistossomose. Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados. Qual seria uma estimativa para a de cura de um doente?

Exercício 4: Numa pesquisa com 50 eleitores, o candidato José obteve 17 votos. Supondo que a eleição fosse hoje, qual seria uma estimativa da proporção de votos para o candidato mencionado?